Highlights

- An urban-scale coupled indoor and outdoor CFD model is developed to study cross ventilation.
- Urban planning and building design features are parametrized and quantified using six key variables.
- A novel ventilation evaluation index is proposed to integrate indoor and outdoor wind environments.
- An advanced machine learning algorithm is used to develop data-driven CIOI prediction model for fast early design support.

Data-driven Model for Cross Ventilation Potential in High-density Cities Based on Coupled CFD Simulation and Machine Learning

Chao Ding[1,2*] and Khee Poh Lam[1,3]

[1]Center for Building Performance and Diagnostics, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

[2]Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, 94720,USA

[3]Department of Building, School of Design and Environment, National University of Singapore, 4 Architecture Dr, 117566, Singapore

[*]Corresponding author: chaoding@lbl.gov

Abstract

Effective urban ventilation through decent urban planning and building design can alleviate the deterioration of the urban built environment. However, natural ventilation requirements and guidelines in current building codes and standards are either qualitative or quantitative but subject to an absolute indoor airspeed threshold without considering the outdoor wind environment. To fill this gap, this paper develops an urban-scale coupled indoor and outdoor computational fluid dynamics (CFD) model and defines a novel ventilation index to assess natural ventilation potential. The index considers wind environments of both indoor and outdoor spaces. First, the coupled CFD model is developed to study wind-driven cross ventilation in high-density cities. A 3D isothermal CFD simulation is solved using the RNG k-ε turbulent model. The simulation results are compared with wind tunnel experiment data from literature. Second, six key design variables are used to generate 3,840 parametric design variations for natural ventilation assessment. Third, a novel integrated index $CIOI_v$ (coupled indoor and outdoor interaction) is proposed to evaluate the wind speed ratio between the indoor area and outdoor reference area. For demonstration, $CIOI_{v,F1}$ is used to represent $CIOI_v$ on the ground level. Data-driven $CIOI_{v,F1}$ models are developed to predict indoor building ventilation potential for quick early design support. Compared with multivariate linear regression model, the Gradient Boosting non-linear model displays much higher prediction accuracy (mean absolute percentage error =0.16 with R^2 =0.8). In early design stage, designers and engineers can skip the computational expensive CFD simulations and use this data-drive model to quickly check the building natural ventilation potentials of different design options in an urban environment.

Keywords: Urban ventilation; Coupled CFD simulation; High-density city; Machine learning; Data-driven model; Early design support

Nomenclature

Symbols		$\bar{v}_{in}$	Indoor weighted average wind speed at z=z_{ref}, [m/s]
k	Turbulent kinetic energy [m^2/s^2]	$\bar{v}_{out,ref}$	Area weighted average speed of the outdoor reference region $A_{out,Ref}$ at z_{ref}, [m/s]
ε	Turbulent dissipation rate [m^2/s^3]		
Q	Flow rate, [kg/s]	Acronyms	
τ	Age of air, [s]	ANN	Artificial Neural Networks
ε_A	Air exchange efficiency, [%]	*SVM*	Support Vector Machine
S_θ	Relative Sinuosity	*VR*	Wind Velocity Ratio
U_{ref}	Velocity at reference height, [m/s]	ACH	Air Change Rate
z_{ref}	Reference height, [m]	HV	Height Variation
z_0	Ground elevation, [m]	WD	Wind Direction
z_G	Roughness length, [m]	OWR	Opening-to-Wall ratio
$\propto$	Pow-law exponent based on urban roughness category	UD	Urban Density
A_P	Building plan area [m^2]	TBH	Target Building Height
A_T	Total land surface area of the lot where the buildings are built, [m^2]	CV	Concave
H_i	Height of building block i, [m]	CX	Convex
H_{ave}	Average height of every building block, [m]	TP	Tapered
N	Total number of building blocks in the computational domain	CIOI	Coupled Indoor and Outdoor Interaction Index
$A_{opening}$	Total opening areas, [m^2]	*MAE*	Mean Absolute Error
A_{wall}	Exterior wall area, [m^2]	*RMSE*	Root Mean Square Error
L_i	Length of the linear segment i, [m]	*NRMSE*	Normalized Root Mean Square Error
θ	Angle between the inflow wind and the linear segment i, [^{0}]	MAPE	Mean Absolute Percentage Error

1. Introduction

The urban population has experienced rapid growth since 1950. By 2050, approximately two-thirds of the population in the world will move to urban areas [1]. An increasing number of high-density cities are formed or will be formed soon. Buildings consume 30~40% of the world's final energy [2,3]. Rapid urbanization, as well as a warming world, requires increasing demand for cooling and building energy consumption. Reducing urban energy use is becoming urgent. As one of the commonly used energy-saving technologies, natural ventilation becomes popular in building designs. Different from mechanical ventilation which uses a mechanical fan to supply and remove air inside a space, natural ventilation relies on natural forces (such as wind or temperature difference) to create airflows in and around buildings. Under good outdoor conditions, efficient natural ventilation can provide better indoor air quality, thermal comfort and productivity for building occupants [4-5]. More importantly, buildings with effective natural ventilation can reduce greenhouse emissions and save 40%-70% of cooling energy compared with mechanical ventilated buildings [6-9]. Therefore,

architects and engineers have started to pay more attention to natural ventilation and prefer operable windows in their designs. Natural ventilation can be further divided into two basic categories: wind-driven ventilation and buoyancy-driven ventilation (or stack ventilation). From a design strategy perspective, natural ventilation includes single-sided ventilation, cross ventilation, and stack ventilation. This paper focuses on wind-driven isothermal cross natural ventilation.

Due to the nature of wind, it is complicated to design natural ventilation. However, computational fluid dynamics (CFD) simulation can provide a feasible way to exam ventilation performance. By solving sets of non-linear governing equations, high resolution temporal and spatial indoor environment information with detailed flow fields can be fully calculated.

As the most efficient natural ventilation design strategy, cross-ventilation can be modeled using two main approaches: the decoupled method and the coupled method (see Fig.1). A decoupled approach needs two sets of CFD models: one for outdoor space and the other for indoor space. Openings are either removed from geometry or modeled as a fully closed surface in the outdoor CFD model. The outdoor airflow field is calculated first. In the second step, an indoor CFD simulation is conducted by mapping the outdoor pressure profiles at the opening positions as boundary conditions for the indoor model. However, many researchers have shown that the decoupled approach has significant errors when applied to large openings [10-13]. These errors come from the assumption that turbulent kinetic energy will vanish at the upwind openings and the presence of the openings will not change pressure distribution on the building envelope. However, experiments have shown that this kinetic energy is still preserved in the indoor space [10-12]. Therefore, recent research uses a coupled approach to simulates cross ventilation through large openings. This approach matches more closely with the actual air flow phenomena because the coupled CFD model contains both indoor and outdoor spaces in the same computational domain. The opening is modeled as a void to connect the external with internal wind environments. Different from the decoupled model which usually assumes that the directions of the mapping vectors are perpendicular to the surface of the openings, the coupled model lets the air to choose the flow path near the openings by itself. The air could either pass through the openings or move around the building.

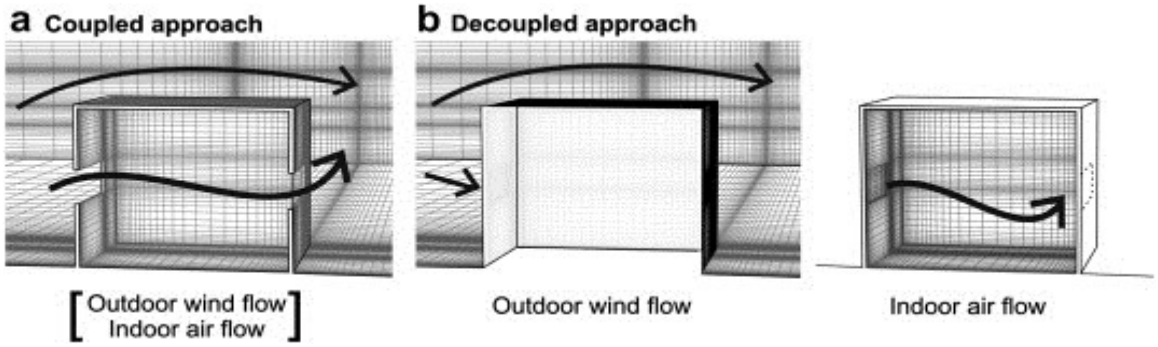


Fig. 1. The coupled and the decoupled approaches [13].

Because of the complexity, researchers typically conducted coupled CFD simulations in a single zone configuration [12-18]. Recently, few papers have studied larger scale coupled indoor and outdoor CFD models with more complex geometries [13, 19-21]. However, they performed either on an isolated building or in a relatively low-density environment. High-density urban areas have

seldom been analyzed. Additionally, high-rise buildings higher than 100 m (such as skyscrapers which are common in high-density cities) are not found in the literature. According to the author's previous research, these buildings display very different features [40].To fill these gaps, this paper conducts a coupled indoor and outdoor CFD model to investigate wind-driven cross ventilation in a high-density urban area.

There are several indices to assess ventilation efficiency and effectiveness in the literature, such as flow rate, air change rate, age of air, air exchange efficiency and wind velocity ratio [15, 19, 22-28]. However, no index is defined to evaluate indoor ventilation performance while considering the outdoor wind environment.

While CFD simulations can provide detailed indoor and outdoor wind patterns, they require a solid engineering background to set up a simulation. Additionally, solving numerical problems can be very expensive in terms of both computational time and hardware requirements. For faster decision making in the early design stage, previous researchers have developed outdoor wind velocity ratio prediction models to forecast outdoor ventilation potential in an urban scale [28-30]. However, theses prediction models are either linear regressions which cannot well describe high-density urban airflow, or non-linear regressions with clear pre-defined equation form which requires strong engineering knowledge and creates a barrier for urban planners and architects. To avoid this problem, more advanced data-driven machine learning algorithm may be a better fit to train a non-linear regression model. Users do not need any assumed equation format. In addition, the source of the data does not have to come solely from simulations. The machine learning model can integrate real data with simulation data to do future prediction.

Although not found in building CFD applications, other non-linear machine learning algorithms, such as artificial neural networks (ANNs) and support vector machines (SVMs) have been widely used in building energy prediction and have shown accurate predictions [31-39].

Table 1 shows a summary of the reviewed literature.

Table 1
Summary of reviewed literatures.

Authors	CFD Model		Configuration		Building Height[m]		Evaluation Parameter					
	External flow	Coupled	Isolated building	Building arrays	<100	>100	Q	ACH	τ	ε	VR	CIOI
Buccolieri et al. (2010) [22]	●			●	●		●		●			
Hang and Li (2010a,b) [23,24]	●			●	●		●	●	●	●		
Lin et al. (2014) [25]	●			●	●		●	●			●	
Ramponi et al. (2015) [26]	●			●	●				●			
Ng (2009) [27]	●			●	●						●	
Lee (2013) [28]	●			●	●						●	
Ignatius et al. (2015) [29]	●			●	●						●	
Yuan et al. (2016) [30]	●			●								
Van Hooff and Blocken (2010) [19]		●			●			●				
Cheung and Liu (2011) [14]		●	●		●							
Larsen et al. (2011) [15]		●	●		●		●					
Ramponi and Blocken (2012) [13]		●	●		●							
Van Hooff and Blocken (2013) [20]		●			●							
Teppner et al. (2014) [17]		●	●		●			●				
Tong et al. (2016) [8]		●	●		●		●					
Shetabivash (2015) [16]		●	●		●		●					
Yang et al. (2015) [21]		●		●	●		●					
Van Hooff et al. (2017) [18]		●	●		●						●	
Kosutova et al. (2019) [61]		●	●		●		●				●	
Arinami et al. (2019) [62]		●	●		●		●					
Liu et al. (2019) [63]		●	●		●		●					
Ikegaya et al. (2019) [64]		●	●		●		●				●	
This paper		●		●	●	●						●*

Note: Q= Flow rate; ACH= Air change rate;τ = Age of air; ε_A = Air exchange efficiency; VR: Wind velocity ratio. *CIOI is a function of 6 key design parameters (Urban Density UD, Height Variation HV, Relative Sinuosity S_θ, Wind Direction WD, Target Building Height TBH and Opening-to-Wall ratio OWR).

2. Base Case Coupled CFD Model Setup

2.1 Computational Domain

Fig. 2 shows the geometry of a generic high-density urban [41]. To reduce computational time with satisfactory accuracy, the computational domain has been resized based on Yoshie's research suggestions [42] and extended followed by the COST Best Practice Guideline [43]. Since the urban geometry is symmetric along y-axis, only half of the computational domain is calculated, and the other half is mirrored in post analysis. The dimensions of the computational domain are 6075 m, 1675 m and 1500 m in x, y and z directions, respectively. There is one skyscraper in the center of the high-density urban area with a dimension of 62.5 m × 62.5 m × 250 m. The high-rise building has 60 floors. The height of each floor is 4 m. Based on DOE (the U.S. Department of Energy) reference buildings' typical building construction, the wall is assumed to be 0.3 m thick [44]. The center of the target building's bottom surface is set to be the origin of the coordinate system. Besides the central skyscraper, there are 48 low-rise building blocks in the computational domain. The dimension of each block is 100 m × 100 m × 25 m.

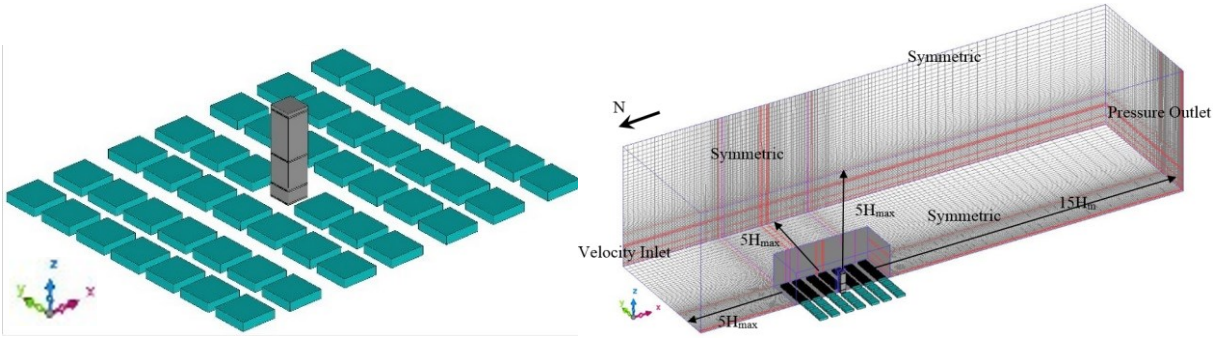


Fig. 2. Geometry, computational domain and boundary conditions.

To achieve a coupled indoor and outdoor CFD model, openings are created on north and south external walls for cross ventilation. In this study, four different floors (F1, F10, F30 and F60) are assumed to have openings. The dimension of the opening is 6 m×2 m×0.3 m. Fig. 3 displays a front view of the opening layout on the first floor. The center of the opening is 2 m above the floor level. It is noted that interior layout, such as furniture, influences the indoor airflow pattern [60]. In this paper, the indoor space is simplified to be empty. In future work, furniture will be considered.

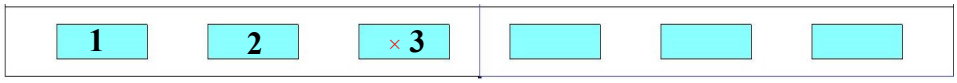


Fig. 3. Illustration of openings on the first floor.

2.2. Boundary Conditions

Align with the incoming wind direction, the front surface of the domain is defined as velocity inlet. The back surface is assigned pressure outlet boundary condition with a zero-gauge pressure. The bottom is set to wall boundary condition. The other three surfaces (top and two sides) are set to be symmetric conditions. The atmospheric boundary condition has been applied to the inlet by defining the vertical profile of velocity, turbulent intensity, turbulent kinetic energy and turbulent dissipation rate based on the AIJ's benchmark tests [42, 45].

$$U(z) = U_{ref}\left(\frac{z-z_0}{z_{ref}-z_0}\right)^{\alpha} \quad (1)$$

$$I(z) = 0.1\left(\frac{z}{z_G}\right)^{-(\alpha+0.05)} \quad (2)$$

$$k(z) \approx [\mathrm{I(z)} \cdot \mathrm{U(z)}]^2 \quad (3)$$

$$\varepsilon(z) \approx P_k(z) \approx C_\mu^{0.5} k(z) \frac{dU(z)}{dz} = \alpha C_\mu^{0.5} k(z) \frac{U_{ref}}{z_{ref}} \left(\frac{z}{z_{ref}}\right)^{\alpha-1} \quad (4)$$

Where U_{ref} is the velocity at reference height, U_{ref}=6.65 m/s in this study according to the inflow velocity measure at the height of the center building by AIJ; z_0 is the ground elevation, z_0=0 m in this study; z_{ref} is the reference height, z_{ref}=250 m (height of the high-rise building); z_G is the roughness length, z_G=450 m according to the AIJ benchmark guidebook; $\propto$ is the pow-law exponent based on urban roughness category, $\propto$=0.20 according to the velocity profile measured from wind tunnel experiment.

2.3. Mesh Grid

To reduce the number of mesh element while preserving accuracy, the all-hex mesh is created for the entire domain. The multi-block method is applied to capture the fluid behaviors. Eight different blocks have been created with different mesh sizes to discretize the domain. Finer mesh for openings and near target building blocks, and coarser mesh for far field. Block 2~8 apply uniform mesh in each block. The Root Block applies a stretching ratio of 1.05. More details can be found in the author's earlier research [46].

2.4. Solver Settings

3D steady state incompressible flow simulations have been performed using scSTREAM v13. RNG k-ε turbulent model is chosen because of its high accuracy [47]. The SIMPLE scheme is used for pressure-velocity coupling. Convergence criteria for continuity, momentum and turbulent (k, epsilon) terms are set to 10^{-5}. Because the first order discretization scheme will bring numerical diffusion, the COST Best Practice Guideline suggests that higher order scheme should be used for the advection approximation [43]. In this paper, the QUICK scheme (3rd order) is applied to advection terms. A double precision solver is used to reduce the round-off error.

2.5. Mesh Setting and Sensitivity Analysis

Mesh sensitivity analysis is performed to find a proper mesh configuration for CFD simulation. Fig. 4 shows four different mesh configurations. Local refinement is implemented near the openings and the target building to improve accuracy. From Mesh1 to Mesh4, the mesh density (standard length) in each mesh block increases by a factor of 0.8, and the number of elements increases approximately by a factor of 2.

To further study the indoor air velocity distribution, a reference line is defined (see the bottom left thumbnail in Fig. 5). The indoor space (62.5 m × 62.5 m) is highlighted in yellow and the outdoor space is colored in blue. The reference line covers both the outdoor and indoor areas and crosses the center of Opening 3 (red '×' in Fig. 3) on Floor 60 at z=238 m. Fig. 5 shows the air speed distributions of different mesh densities along the reference line. The horizontal axis represents the relative location of the space in x-direction (x) divided by the total length of the target building (L). In other words, the front opening is at x/L = 0 while the back opening is at x/L= 1. x/L < 0 indicates upwind region and x/L>1 is downwind region. The

simulation discrepancy between Mesh3 and Mesh4 are minor. To balance computational time and simulation accuracy, the following analysis applies Mesh3.

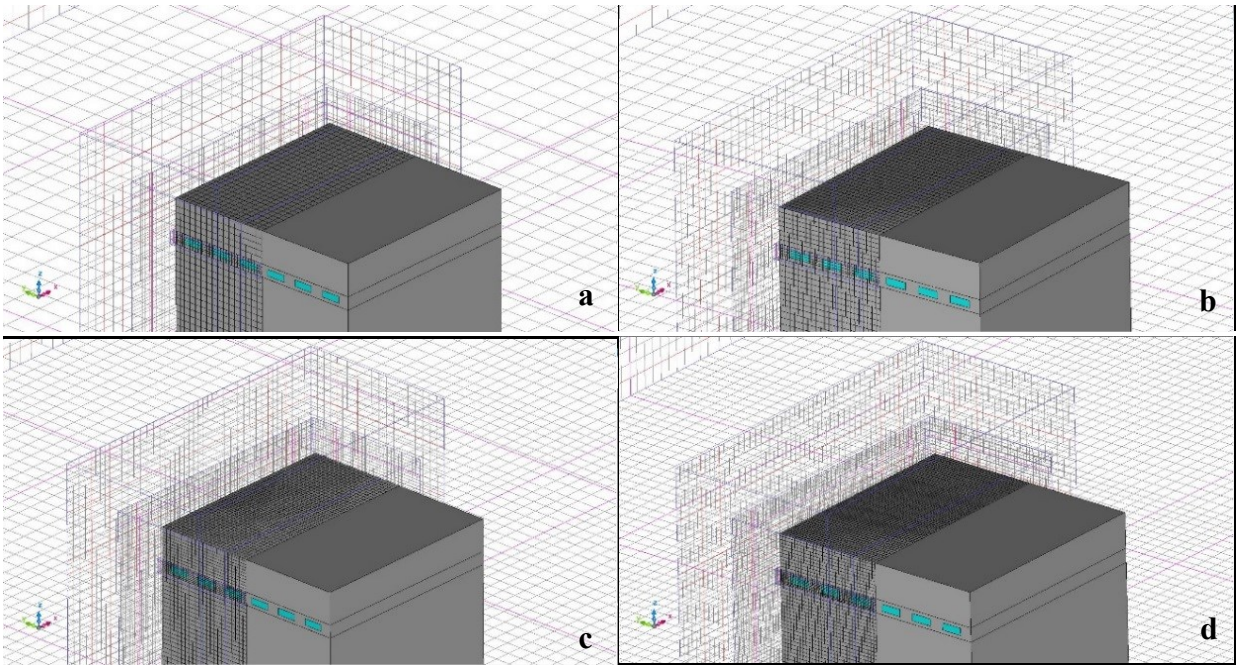


Fig. 4. Different mesh configurations - (a) Mesh1; (b) Mesh2; (c) Mesh3 and (d) Mesh4.

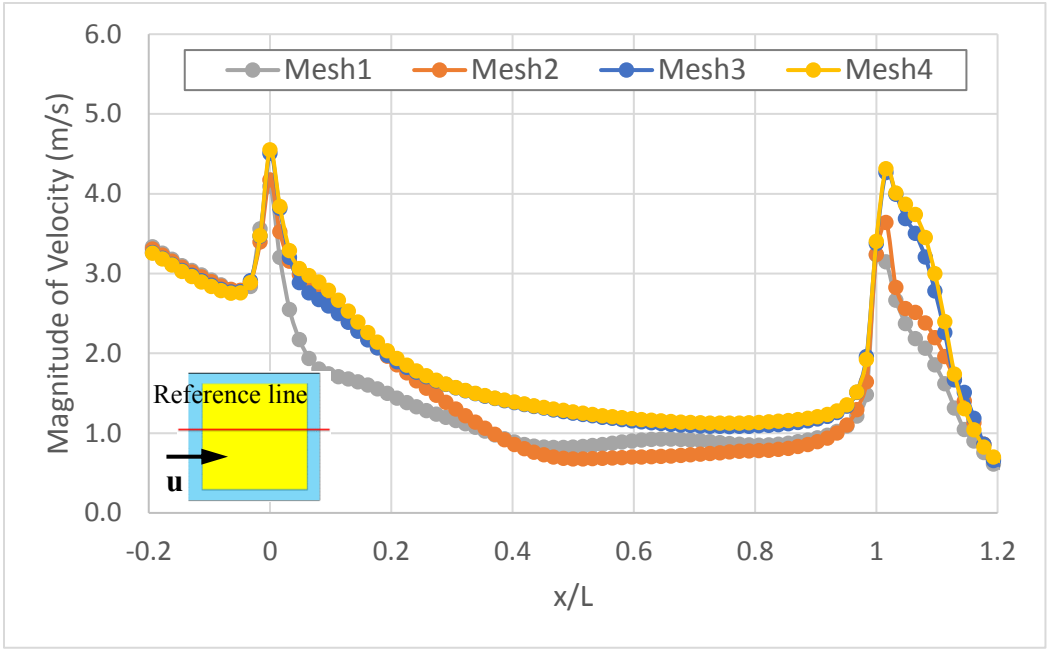


Fig. 5. Mesh sensitivity analysis (z=238m, red line in the thumbnail indicating reference line).

2.6. Result Comparison with Wind Tunnel Experiment

Mesh3 simulation results are compared with wind tunnel experiment from literature [42]. Locations of different measuring points can be found in Fig.6. Fig. 7 shows the comparison between Mesh3 simulation results (blue square) and wind tunnel measurement (green circle). Length of each straight line indicates the magnitude of difference at each measuring point. Sufficient agreement in the external wind flow field has

been achieved. Mean Absolute Percentage Error (MAPE): $MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$ *100%=0.28, R^2_{adj}=0.78. This overall performance is acceptable for urban-scale simulations.

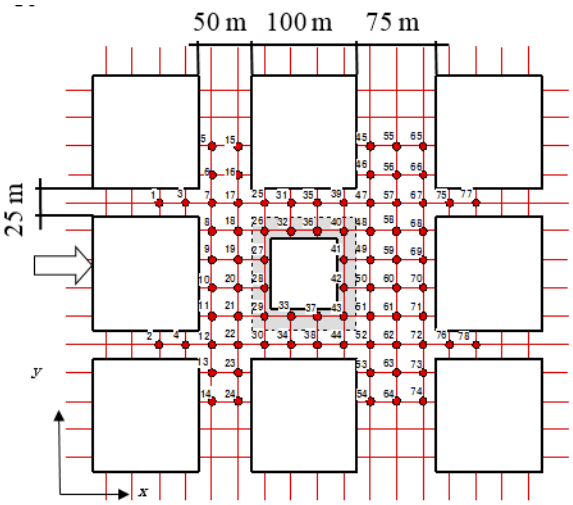


Fig.6. Measuring points [42].

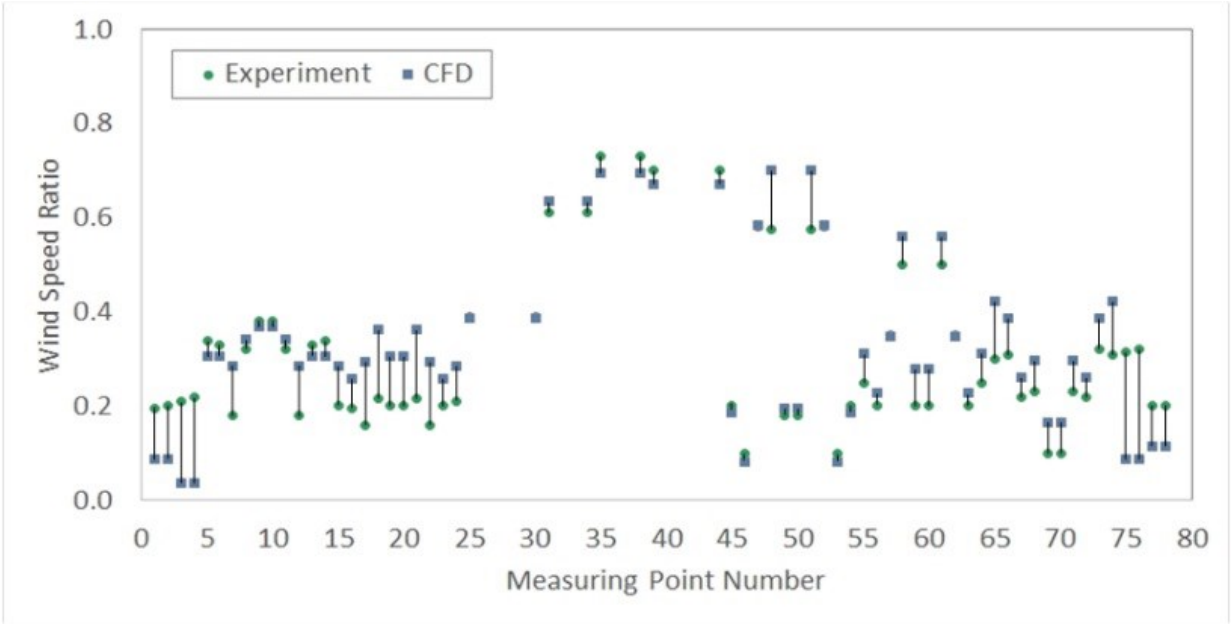


Fig. 7. Comparison between CFD and experiment.

2.7. Result Analysis of Base Case Model

Fig. 8 shows the air speed distribution at z = 2m (1st Floor). Fig. 8 (a) indicates that the outdoor airflow is fully developed in the extended computational domain. No backflow is observed near the outlet boundary. Fig. 8 (b) is a magnified view of the target building and its adjacent outdoor region. It can be clearly detected that outdoor wind does enter the room from openings of the first floor. The indoor air distribution is influenced by the local microclimate outside of the target building. Low-velocity region is observed in the upwind region. High-velocity regions are found when air hits sharp edges, such as corners of the exterior

walls in the lateral areas of the target building. Inside the room, velocity contour shows clear cross ventilation pattern with a relatively higher velocity near openings and lower velocity in the center.

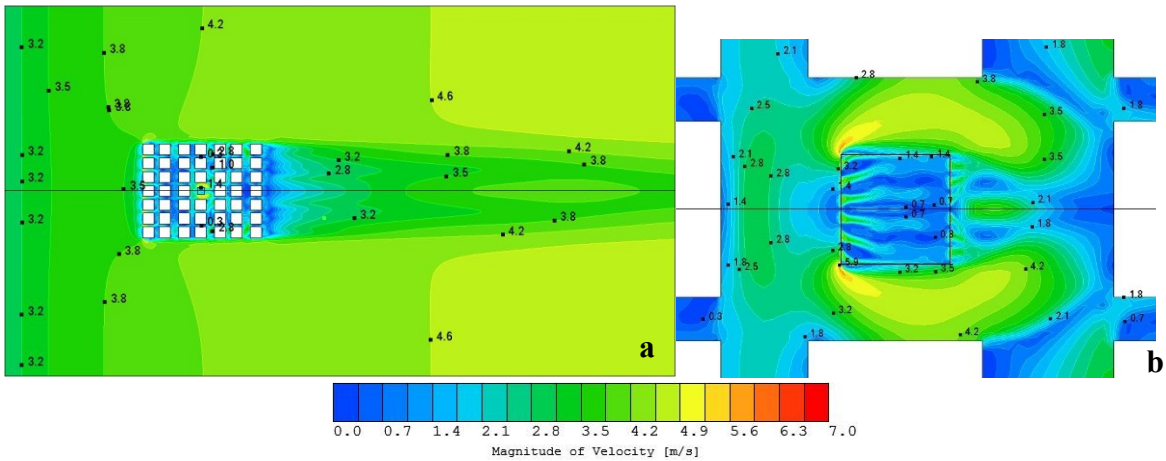


Fig. 8. Air speed distributions at z=2 m (1st Floor) (a) entire domain; (b) magnified view.

3. Parametric Study

Stochastic simulation is conducted to generate various urban and building design variations. Researchers have shown that outdoor wind environment can be affected by several morphological variables [48-53]. Based on current literature and the authors' expertise, six key design parameters are used to digitalize and quantify important urban planning and building design features, which cover outdoor urban morphology, building design and wind meteorology.

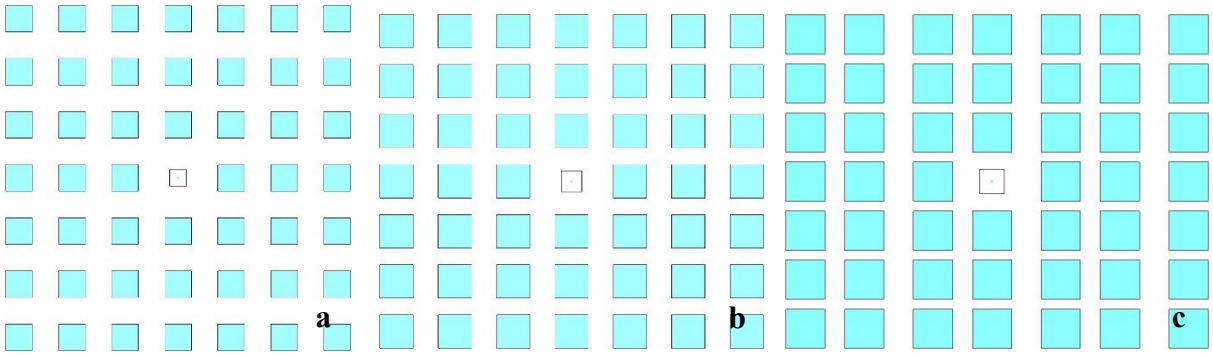
3.1. Urban Density (UD)

Building density or plan area density can be used to describe urban density [22, 26, 54-55]. Grimmond and Oke defined λ_P as shown in the following equation.

$$\lambda_P = {}^{A_P}/_{A_T} \quad (5)$$

Where A_P is the building plan area; A_T is the total land surface area of the lot where the buildings are built.

Four different urban densities (0.29, 0.42, 0.53, 0.70 and 0.75) are created for parametric study by changing urban street width to generate different λ_P (see Fig. 9). The size of the computational domain for each urban density scenario is adjusted to ensure the flow field is fully developed in the domain.



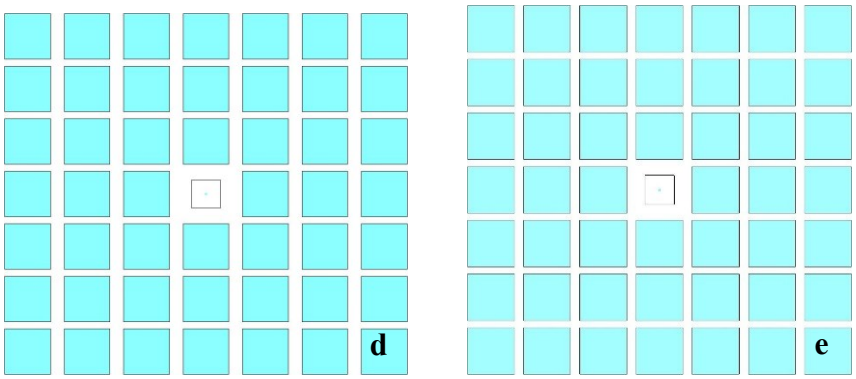


Fig. 9. Urban density scenarios (a) 0.29; (b) 0.42; (c) 0.53; (d) 0.70; (e) 0.75.

3.2. Target Building Heights (TBH)

Target building height (TBH) is defined as the height of the target building. Four target building heights (25m, 50m, 100 m, and 250m) are discussed in this section. Fig. 10 demonstrates the four different TBHs. Similar to the base case model, openings are created on three different floors: bottom floor, middle floor and top floor (TBH250 has four floors with openings).

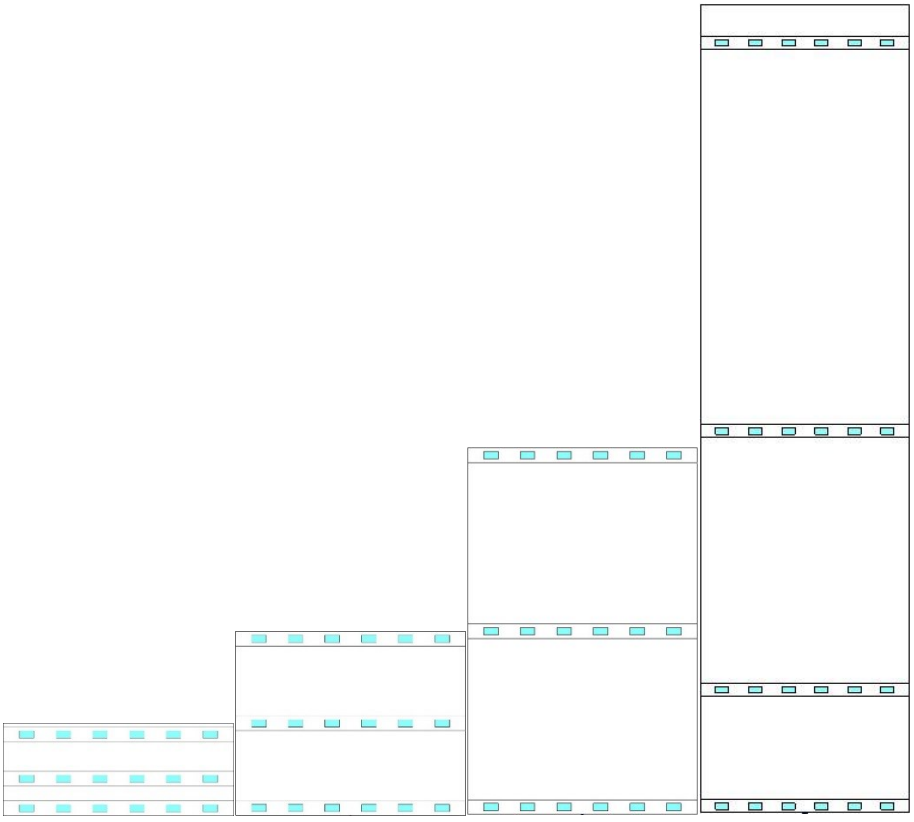


Fig. 10. Four target building heights (from left to right: 25m, 50m, 100m and 250m).

3.3. Buildings' Height Variation (HV)

3.3.1. Height Variation (HV)

Height variation (HV) of each scenario is computed to quantify the building height distribution in the urban area using Equation 6.

$$HV = \sqrt{\sum_{i=1}^{N}(H_i - H_{ave})^2 / N} \quad (6)$$

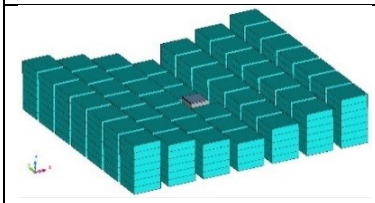
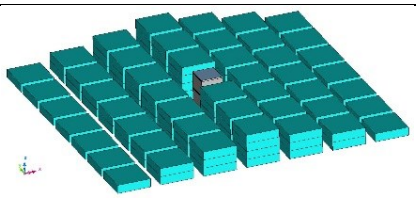
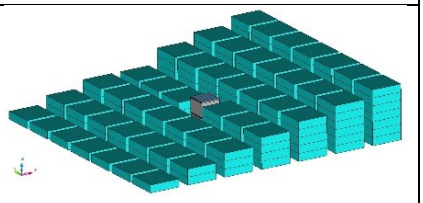
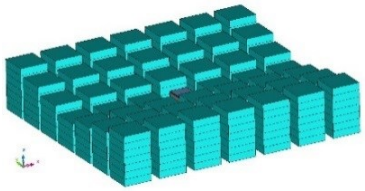
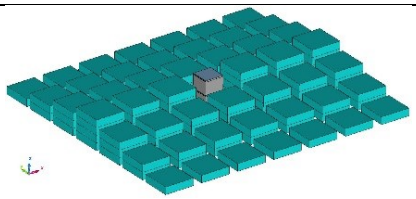
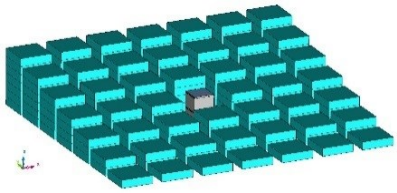
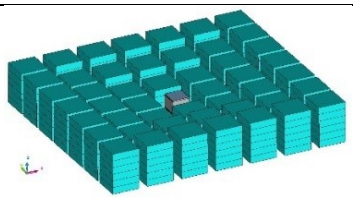
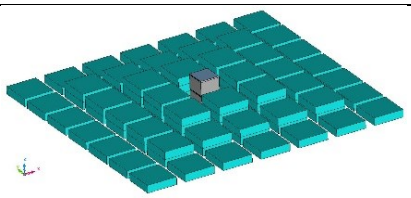
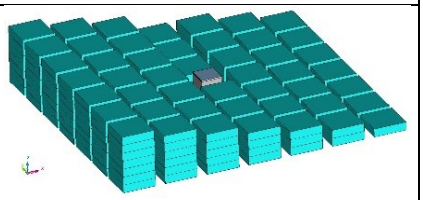
Where H_i is the height of building block i, [m]; H_{ave} is the average height of every building block, [m]; N is the total number of building blocks in the computational domain

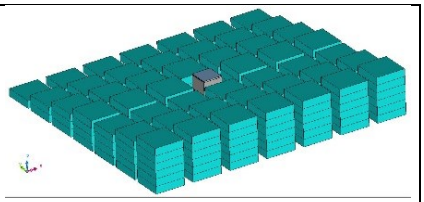
It can be observed that HV is the standard deviation of building height H_i. For a uniform urban configuration, HV=0, which is not recommended from the ventilation point of view. The unit of HV is m.

3.3.2. Urban Form

To generate different HV values, three urban archetypes are defined. They are Concave, Convex and Tapered, which are commonly seen in urban planning. Concave (CV) refers to higher outer building blocks and lower inner building blocks (such as Beijing); Convex (CX) refers to higher inner and lower outer blocks (such as the Tower Bridge area in London); and Tapered (TP) refers to building block heights increase/decrease towards one end (such as some areas in San Francisco). Since the base case urban geometry has symmetric feature, in each type of the main urban form, sub-modules are created by rotating the CX, CV and TP certain degrees along z-axis. Using this rotation transformation not only remove the impact from the symmetric characteristic of the base case, but also modularize the urban form. With different individual urban form modules, it is possible to generate more scenarios by combining corresponding modules.

Table 2
Urban form modules.

Concave	Convex	Tapered
		
		
		

3.4. Opening Size

In coupled CFD models, indoor and outdoor environments are connected through openings. The opening design will highly influence indoor air distribution. In this paper, all openings are assumed to be fully open to achieve maximum cross ventilation potential. Four different window sizes are considered to understand the influence of window design. Various scenarios are created by keeping the height of the opening constant while increasing the opening length. The minimum opening size is 2 m × 4 m (height × length) and the maximum opening is full glazing (2 m × 58 m). Fig. 11 shows the four different opening configurations when the target building height is 50 m.

Opening-to-Wall Ratio (OWR) is used to convert opening size to a dimensionless parameter for further analysis. OWR can be calculated using Equation 7.

$$\mathrm{OWR} = {}^{\mathrm{A_{opening}}}\!/_{\mathrm{A_{wall}}} \tag{7}$$

Where $\mathrm{A_{opening}}$ is the total opening areas $[m^2]$; $\mathrm{A_{wall}}$ is the exterior wall area where the openings are located $[m^2]$

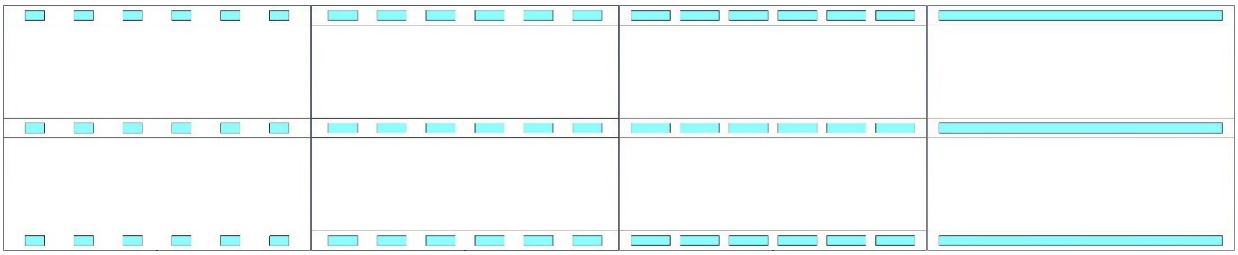


Fig. 11. Four different opening sizes (TBH=50m) (a) 2 m × 4 m; (b) 2 m × 6 m; (c) 2 m × 8 m and; (d) 2 m × 58 m.

Openings are created on different levels (bottom level, middle level and top level) of the target building, which provides the feasibility to study indoor flow behaviors of different height. With more individual access, the bottom floor is more people orientated while the top floor is more building structure orientated. Although each floor has its own unique airflow pattern, changing cross ventilation floors will not change the nature of the entire research process as proposed in this chapter as well as the followings. This paper uses the 1st floor as a demonstration.

3.5. Wind Direction (WD) and Orientation of Urban Canyons

3.5.1. Wind Direction

Wind direction plays an important role in urban flow patterns. By changing different wind directions, indoor airflow field varies a lot even applying the same inlet boundary conditions. More details can be found in the author's other paper [46]. Four different wind directions (0^0, 22.5^0, 45^0, 67.5^0) are considered in this section. 0^0 is defined as North wind direction. Consequently, 45^0 means NE wind.

3.5.2. Relative Sinuosity (S_θ)

Besides wind direction, Orientation of Urban Canyons is another significant parameter related to directional information. Relative Sinuosity (S_θ) is used to parameterize and quantify this variable using Equation 8. S_θ measures the influence of wind penetration through street canyons under each wind direction [48].

$$S_\theta = \sum L_i \cos^2(\theta_i) \Big/ \sum L_i \tag{8}$$

Where L_i is the length of the linear segment i, [m]; θ is the angle between the inflow wind and the linear segment i, [^{0}]

Fig. 12 is a plan view of the generic urban model. Black arrows on the top left corners indicate four different wind directions defined above. The red double arrows represent urban canyons without any obstacles, also known as free breezeways. The incoming wind can freely go through the entire urban area from one side to another and creates accelerated wind speed in these breezeways. The length of the horizontal breezeway is assumed to be L_1 and the length of the vertical one is assumed to be L_2. In this figure, there are 6 horizontal and 6 vertical linear segments (breezeways), respectively. θ in Equation 14 is the angle between the inflow wind direction and the corresponding breezeway. For example, use θ_i to represent the angle between inflow wind direction and horizontal breezeways and use $\theta_i'$to present the angle between inflow wind direction and vertical breezeways. Then $\theta_i + \theta_i'$=90^0. If $\theta = 0^\circ$ (north/south direction), $\theta' = 90^\circ$. it means that the incoming wind is parallel to the horizontal breezeways (max impact) and vertical to the vertical breezeways (no impact). When $0^\circ < \theta_i < 90^\circ$, both parallel and horizontal impact are included. S_θ considers the acceleration projection onto vertical and horizontal axes.

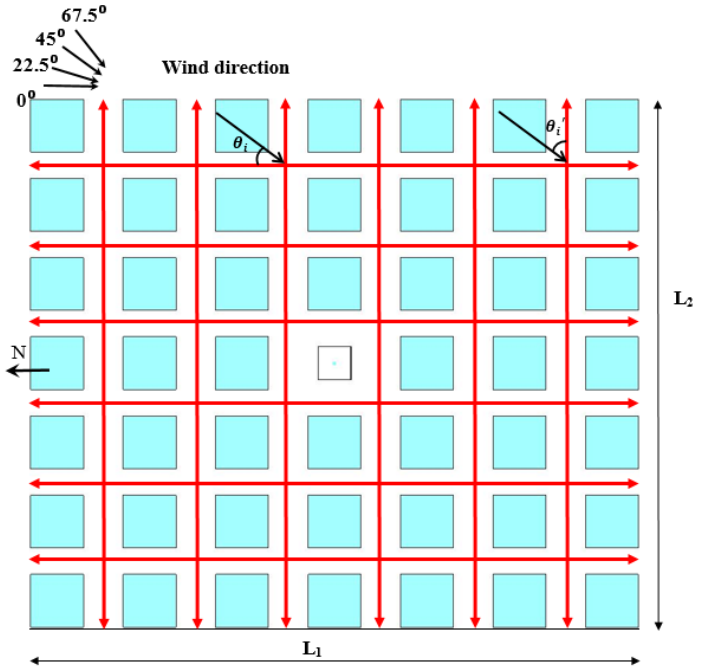


Fig. 12. Plan view of generic urban model.

3.6. Summary of Parametric Scenarios

Fig. 13 shows the total 1,840 CFD parametric combinations. Each design scenario is digitalized and characterized by its six key design parameters. Thanks to the symmetric feature of the base case urban geometry, combined with different urban form modules, the numerical CFD database can be expanded from the original 4 wind directions to all 16 wind directions. Table 3 shows the formation of the 16 different wind conditions through re-grouping and combining certain urban form modules. For example, N0 wind condition is formed by combining CX_X, CV_X and TP_0X urban form modules of 0^0 wind direction. Similarly, N225 wind conditions are formed by combining CX_X, CV_X and TP_180X urban form modules under 45^0 wind direction. Following this rule, the numerical CFD database is expanded to 3,840 design scenarios to cover 16 wind directions.

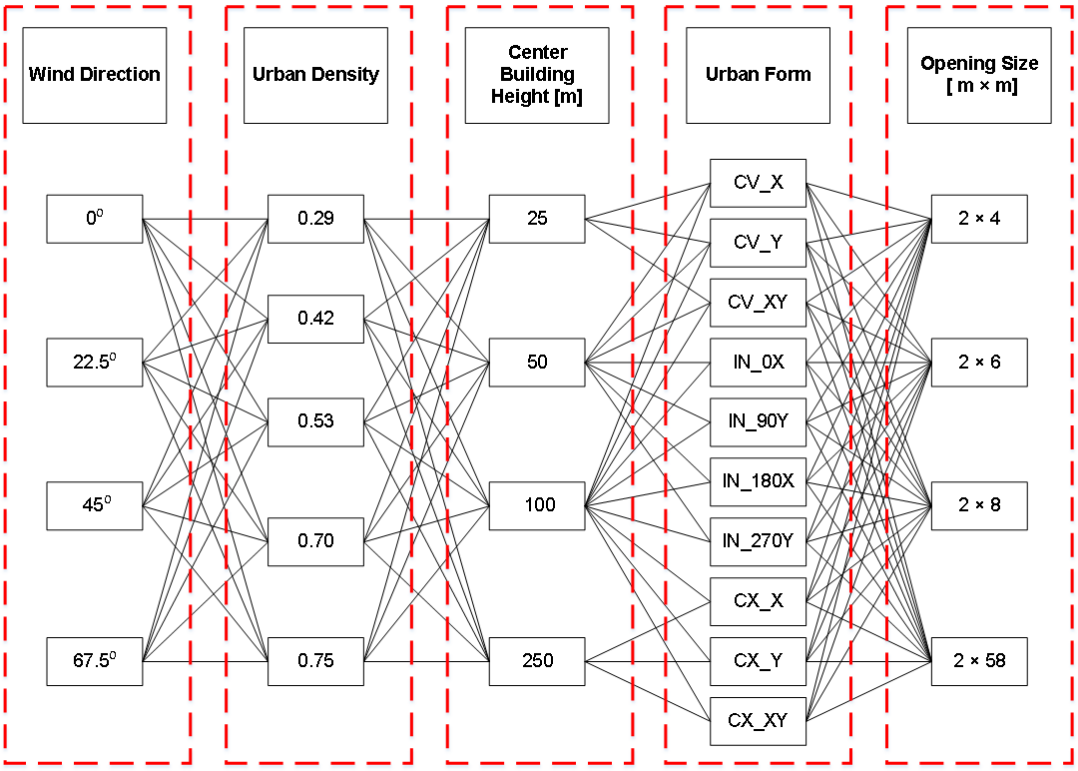


Fig. 13. Combination of 1,840 possible design scenarios.

Table 3
Formation of 16 wind direction cases.

Wind Direction	Rotation Base	Rotation Angle from N (^{0})	Urban Form
N0	-	-	CX_X/CV_X/TP_0X
N22.5	-	-	CX_X/CV_X/TP_0X
N45	-	-	CX_X/CV_X/TP_0X
N67.5	-	-	CX_X/CV_X/TP_0X
N90	N0	90	CX_Y/CV_Y/TP_90Y
N180	N0	180	CX_X/CV_X/TP_180X
N270	N0	270	CX_Y/CV_Y/TP_270Y
N112.5	N22.5	90	CX_Y/CV_Y/TP_90Y

N202.5	N22.5	180	CX_X/CV_X/TP_180X
N292.5	N22.5	270	CX_Y/CV_Y/TP_270Y
N135	N45	90	CX_Y/CV_Y/TP_90Y
N225	N45	180	CX_X/CV_X/TP_180X
N315	N45	270	CX_Y/CV_Y/TP_270Y
N157.5	N67.5	90	CX_Y/CV_Y/TP_90Y
N247.5	N67.5	180	CX_X/CV_X/TP_180X
N337.5	N67.5	270	CX_Y/CV_Y/TP_270Y

4. CIOI Prediction Models

4.1. Definition of CIOI Index

The definition of CIOI index is given in Equation 9. CIOI is a dimensionless index, which integrates both indoor and outdoor wind environments.

$$\mathrm{CIOI} = \frac{\Phi_{in}(x,y,z)}{\Phi_{out,ref}} \quad (9)$$

where $\Phi_{in}(x, y, z)$ is an indoor parameter (such as velocity, pressure or pollutant concentration) at location (x,y,z) in space; $\Phi_{out,ref}$ is an average value at a pre-defined outdoor region to represent the reference outdoor wind condition near the building

CIOI is used to assess indoor cross ventilation potential while considering outdoor wind environment. Indoor layouts (such as furniture layout) are ignored. The indoor space is assumed to be an open space with no indoor obstacles. This assumption could be applied to office buildings.

Since the defined CIOI index is generic, this paper uses wind velocity as an example to demonstrate the implementation of the CIOI concept and methodology. The same procedure can be used for CIOI indices of other air variables, such as pressure, temperature and pollution concentration. To demonstrate the velocity prediction, the reference height is assigned to the same height as the opening center. The magnitudes of the area weighted average wind velocity is calculated using Equation 10 to represent wind fields of both indoor and outdoor.

$$\mathrm{CIOI}_v(\mathrm{z} = \mathrm{z}_{ref}) = \frac{\bar{v}_{in}}{\bar{v}_{out,Ref}} \quad (10)$$

where z_{ref} is the same height as the opening center, m; $\bar{v}_{in}$ is the indoor weighted average wind speed at z=z_{ref}, m/s; $\bar{v}_{out,ref}$ is an area weighted average speed of the outdoor reference region $A_{out,Ref}$ at z_{ref}, m/s.

Since openings were created on different floors of the target building, the generic urban CFD database includes indoor and outdoor airflow information of multiple levels (ground level, middle level, and top level). Results of the ground level are used to demonstrate CIOI applications in this paper. It is noted that all the following results and conclusions are based on the results from ground level, which may not apply to different levels. $\mathrm{CIOI}_{v,F1}$ is used to represent the 1st floor (ground level) CIOI_v values.

Fig. 14 shows an example of $\mathrm{CIOI}_{v,F1}$ calculation on the 1st floor when urban density λ_p=0.75. This figure shows a screenshot of z_{ref}=2m, which goes across the center of the 1st floor openings. The indoor area of

the target building is colored in yellow. The area weighted average speed is given by $\bar{v}_{in} = \frac{\frac{1}{N_1}\sum_1^{N_1} v(x,y,2)}{A_{indoor}}$, where N_1 is the number of indoor (yellow region) grid points. The building blocks in the outdoor environment are colored in blue. There are eight surrounding blocks around the target building in this figure. The green part is the reference outdoor region. $A_{out,Ref}$ is defined as the adjacent outdoor reference area of the target building. Thus, the area-weighted average air speed of the outdoor reference area is calculated using $\bar{v}_{out,ref} = \frac{\frac{1}{N_2}\sum_1^{N_2} v(x,y,2)}{A_{out,Ref}}$, where N_2 is the number of reference grid points in the green region.

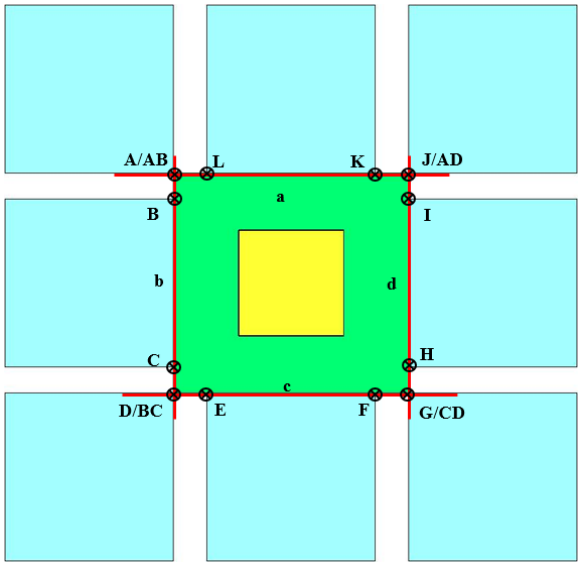


Fig. 14. An example of CIOI calculation at Floor 1 (λ_p=0.75 case).

To obtain the boundary of $A_{out,Ref}$ (see Fig. 14)

1. Draw lines along the edge of each adjacent building block (see red lines a, b, c and d), use AB, BC, CD and DA to represent the intersection of a, b, c and d;
2. Find the corner of each adjacent building (A, B, C, D, E, F, G, H, I, J, K and L);
3. Connect all the points in 1 and 2 in sequence to form a polygon ABCDEFGHIJKLA
4. $A_{out,Ref} = A_{\text{ABCDEFGHIJKLA}}$

From Equation 10, CIOI_v can be used to evaluate cross ventilation potential to support early design. When $0<\text{CIOI}_v<1$, a larger CIOI_v usually indicates a smaller velocity loss after the air enters the indoor space through openings, which shows higher building natural ventilation potential. After calculating the CIOI_v, designers can compare the indices of different design scenarios. However, it is possible to obtain a $\text{CIOI}_v>1$, when a backflow or vortex is generated in the indoor space due to high-density urban morphology and complex coupled wind fields of indoor and outdoor. From a ventilation perspective, outdoor stagnant air, exhaust heat, or pollution may be brought back with the backflow into indoor space under this circumstance, which is bad news for the indoor environment. Therefore, $\text{CIOI}_v > 1$ indicates poor ventilation design, which should be avoided by the designer.

4.2. Prediction Model Development

Coupled CFD simulations can provide detailed indoor and outdoor wind patterns. However, it requires a solid engineering background to set up a simulation and it is very expensive to solve the problem with computers both in terms of computational time and hardware requirement, which creates barriers for designers to use. To make quick decision support available in the early design stage, machine learning models are generated to predict CIOI_v based on stochastic CFD numerical simulations.

In each case, $\bar{v}_{in}$ and $\bar{v}_{out,Ref}$ values of every cross-ventilation floor (first floor, middle floor and top floor) need to be calculated to compute CIOI_v. VB scripts are developed to automatically perform post-analysis in CFD Postprocessor, calculate and produce the estimation of $\bar{v}_{in}$, $\bar{v}_{out,Ref}$ and $CIOI_v$. The main processes are as follows:

1. Launch Postprocessor and read each CFD result (.fld file)
2. Create neutral file objects for the specified region $A_{in}$and $A_{out,Ref}$ of each cross-ventilation floor
3. Selects a scalar variable and specify variable: Magnitude of Velocity (VECTV)
4. Execute scalar integration to calculate $\bar{v}_{in}$ and $\bar{v}_{out,ref}$
5. Save the results to a .csv file
6. Quit Postprocessor and clean up memory

After all the $\bar{v}_{in}$ and $\bar{v}_{out,ref}$ are calculated, another Java code is used to 1. collect .csv files, 2. read $\bar{v}_{in}$ and $\bar{v}_{out,ref}$ values, 3. compute $CIOI_v$ and 4. fill the values into a pre-define excel template.

For the 3,840 CFD simulation cases, CIOI values of the first floor are calculated and collected into a csv file using a Java code to demonstrate the process of prediction model development. Each CFD case generates a list including the six key design variables and one $\mathrm{CIOI}_{v,F1}$ value from Equation 10. The output format is (WD, Se, λ_P, TBH, HV, OWR,$CIOI_{v,F1}$), which is a 7-tuples. All the 3,840 7-tuples are shuffled and divided into two subsets for model development and testing: 70% (2,688 CFD cases) for cross-validation and 30% (1,152 CFD cases) for final test. For each 7-tuples, the 6 design variables are treated as features to generate an estimated $CIOI_{v,F1}{}'$ using the prediction models. Difference between estimation from prediction models and CFD results is analyzed and compared to select the appropriable regression algorithm.

4.2.1. Error Measures

The following metrics are used to evaluated statistical performance of regression models.

- Mean Absolute Error (MAE): $MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$ (11)
- Root Mean Square Error (RMSE): $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$ (12)
- Normalized Root Mean Square Error (NRMSE): $NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}}{y_{max} - y_{min}}$ (13)
- Mean Absolute Percentage Error (MAPE): $MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| * 100\%$ (14)

Where y_i = true value (CFD simulation results); y_{max}=the maximum values (CFD simulation results); y_{min}=the minimum values (CFD simulation results); $\hat{y}_i$= predicted value (regression model outputs)

4.2.2. Multivariate Linear regression (MLR) Model

Multivariate Linear Regression (MLR) is the simplest and the most broadly utilized regression model for prediction. LR draws a line through a multidimensional space. Equation 15 shows the MLR model.

$$F(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (15)$$

Where α is a constant term; $\beta_0 \sim \beta_n$ are coefficients for independent variables; $x_0 \sim x_n$ are all independent variables, in this case, n=6

LR is based on a strong independence assumption. It assumes that each input variable X is independent of each other. There is no or very little multicollinearity in the data set. LR learns and assigns different coefficients (weights) to different independent variables in the model. The total squared errors between the prediction and the actual value are calculated using Equation 16. Through model training, LR aims to minimize the error iteratively by varying the weights of each independent variable X until an optimal solution is obtained.

$$Error = \sum_{i=1}^{n} \left(F(x)_{predict,i} - F(x)_{actual,i}\right)^2 \quad (16)$$

Multivariable linear regression analysis is conducted using the stepwise method. The linear equation is shown below.

$$CIOI_{F1} = -0.000101WD + 0.876S_\theta + 0.2080\lambda_P + 0.000547CBH - 0.001908HV + 0.9793WWR - 0.43 \quad (17)$$

Table 4 shows the performances statistics on coefficients of the training set. It can be noticed that WD and HV have negative correlations on $CIOI_v$ while other design variables have positive impacts. In addition, the variance inflation factor (VIF) quantifies the severity of multicollinearity. VIF=1 means not related, while 1<VIF<5 means moderately correlated. If VIF>10, then multicollinearity is high. We can see that the VIF values of each design variable are less than 2, which means the independence of each variable is acceptable. The ANOVA results of all 2,688 CFD simulation cases in the training set can be found in

Table 5.

Table 4
Linear regression results of all 2,688 CFD simulation cases in training set.

Term	Coefficient	Standard Error Coefficient	95% Confidence Interval	T-value	P-value	VIF
Constant	-0.438	0.0753	(-0.5856, -0.2904	-5.82	0	
WD	-0.0001	0.00003	(-0.000164, -0.000038)	-3.13	0.002	1.02
S_θ	0.876	0.145	(0.592, 1.160)	6.05	0	1.02
λ_P	0.208	0.0195	(0.1699, 0.2462)	10.69	0	1
TBH	0.0005	0.00006	(0.000430, 0.000665)	9.12	0	1.7
HV	-0.0019	0.00038	(-0.002648, -0.001167)	-5.05	0	1.7

OWR	0.9793	0.033	(0.9147, 1.0440)	29.71	0	1

Table 5
ANOVA results of all 2,688 CFD simulation cases in training set.

Source	Sequential sums of squares	Contribution	Adjusted sums of squares	Adjusted mean squares	F-value	P-value
Regression	33.581	0.2953	33.5805	5.5968	187.27	0
WD	0.637	0.0056	0.2931	0.2931	981%	0.002
S_θ	1.198	0.0105	1.0936	1.0936	36.59	0
λ_P	3.092	0.0272	3.4167	3.4167	114.33	0
TBH	1.555	0.0137	2.4834	2.4834	83.1	0
HV	0.724	0.0064	0.7631	0.7631	25.53	0
OWR	26.375	0.232	26.3754	26.3754	882.55	0
Error	80.123	0.7047	80.1226	0.0299		
Total	113.703	1	112.090			

From the above, although the six design variables (WD, S_θ, λ_P, TBH, HV, OWR) show statistical significance ($p<0.05$) on $CIOI_v$ index and the independence assumption of MLR is valid, the linear prediction can only predict approximately 29% of $CIOI_v$ index (See Table 6). The performance of this MLR is tested on the test set (30%, 1,152 CFD cases). We can see that the adjusted R^2 on the test set is similar to that on the training set. The overall MLR prediction performance is very limited. One of the most important metrics for this problem, MAPE, is only 0.5 for the test set, indicating that only about half of the MLR $CIOI_v$ prediction matches with the CFD simulation results. The data cannot be fit into a linear model.

Table 6
Statistics of LR model on Training set and Test set.

Data Set	Training Set (70%)	Test Set (30%)
R^2	0.30	0.29
R^2 (adj)	0.29	0.29
RMSE	0.17	0.18
NRMSE	0.13	0.13
MAE	0.14	0.14
MAPE	0.49	0.50

Since the $CIOI_v$ index values are for high-density ($\lambda_P \geq 0.29$) urban areas, the outdoor flow pattern is complex due to small gaps among building blocks. The integration of flow interactions between indoor and outdoor makes $CIOI_v$ index even more complicated. It is not surprising that linear regression cannot provide a satisfied prediction. The relationship between design variables and $CIOI_v$ index may be higher order polynomial, exponential, power or any other nonlinear format, which MLR is not powerful enough to capture the relationship.

4.2.3. Non-linear Regression Model

Due to the complicated coupled flow behavior in a high-density urban area, the nonlinear format between design variables and $CIOI_v$ is not clear. To provide a good prediction without knowing the model formulation, Gradient Boosting algorithms is selected for further research.

As one of the most powerful machine learning algorithms, Gradient Boosting (GB) is widely used for prediction model development. During the 2015 Kaggle competition, a competition like the Formula Racing for data science, 17 out of the 29 challenge winning solutions used GB [56]. GB trains multiple models and then combines the predictions. It uses an iterative method to integrate several weak classifiers into a strong classifier. The more iterations, the more confident the trained classifier becomes in its predictions.

This paper uses a GB Regressor [57] to implement the gradient boosting of regression tree proposed by Friedman [58]. Equation 18 shows the formulation of the k iterations.

$$F_k(x) = T_1(x) + T_2(x) + \cdots + T_k(x) \tag{18}$$

where $T_k(x)$ is a regression tree

The GB model in the next iteration will be updated by adding a new regression tree $T_{k+1}(x)$ to the previous $F_k(x)$ model (see Equation 19):

$$F_{k+1}(\mathrm{x}) = T_1(x) + T_2(x) + \cdots + T_k(x) + T_{k+1}(x) \tag{19}$$

$T_{k+1}(x)$ is trained to minimize the difference between the true value y and the $F_k(x)$ prediction.

More iterations may reduce the MSE to almost 0 and increase the correlation (R) to nearly 1 on the training set. However, the model may be overfitted and lead to an even worse prediction on the test set. To avoid overfitting, early stopping rule can provide guidance to decide the number of iterations for GB model training. Fig. 15 is an illustration of the early stopping principle based on cross-validation data. The training should be suspended as soon as the error on the cross-validation set is larger than it was of the last iteration.

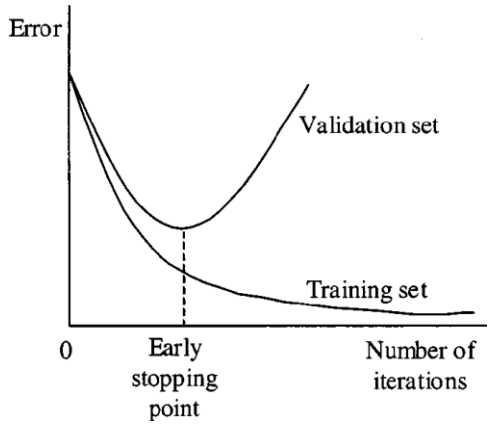


Fig. 15. Early stopping based on cross-validation [59].

The multicollinearity requirement of each design variable in the linear regression model has been checked. Since the GB model is more robust in multicollinearity situations than linear regression, the algorithm itself will handle any correlation among input variable Xs. In each regression tree, many piecewise constant functions are implemented to approximate y values smoothly. Theoretically, the GB model is able to estimate the input variables in any combination forms (i.e. quadratic, logarithmic, exponential, trigonometric, etc.) and provide a reliable prediction.

There are two main parameters can be tuned in the GB model. The first one is the number of training iteration (i) and the other one is the number of leaves in each tree (n). The early stopping principle is applied to control the i value. When the test set reaches the minimum RMSE, the model training is completed. To tune parameter n, the entire data set is divided into 70% cross-validation set and 30% final test set. 5 train/test pairs are created to conduct a 5-fold cross validation. Data in each fold is further divided into a training set (80%) and a testing set (20%) to train and test the model. Table 7 shows the results of the 5-fold cross validation. It can be concluded that when n equals to 20, all statistical metrics show the best results in terms of lowest errors (RMSE, NRMSE, MAE and MAPE) and highest correlation (R^2_{adj}). Thus, n=20 is used to train the final model.

Table 7
Parameter tuning on GB to determine n value for generic urban case.

n	Metrics	Fold1	Fold2	Fold3	Fold4	Fold5	Average
n=10	R^2_{adj}	0.85	0.88	0.88	0.86	0.89	0.87
	RMSE	0.10	0.10	0.10	0.11	0.10	0.10
	NRMSE	0.10	0.08	0.08	0.09	0.07	0.08
	MAE	0.07	0.07	0.07	0.08	0.07	0.07
	MAPE	0.23	0.19	0.22	0.22	0.19	0.21
n=15	R^2_{adj}	0.86	0.88	0.89	0.87	0.89	0.88
	RMSE	0.10	0.10	0.10	0.11	0.10	0.10
	NRMSE	0.10	0.08	0.08	0.09	0.07	0.08
	MAE	0.07	0.07	0.07	0.07	0.07	0.07
	MAPE	0.23	0.19	0.20	0.21	0.19	0.20
n=20	R^2_{adj}	0.85	0.88	0.88	0.86	0.90	0.87
	RMSE	0.10	0.10	0.10	0.11	0.09	0.10
	NRMSE	0.10	0.08	0.08	0.09	0.07	0.08
	MAE	0.07	0.07	0.07	0.08	0.07	0.07
	MAPE	0.22	0.19	0.20	0.21	0.18	0.20

Finally, the GB model stops at i=3980 iterations. The statistical performance on the final test set is summarized in Table 8. Compared with MLR, GB model has achieved much lower error measures with a much higher R^2_{adj} g. One of the most important metrics for our prediction purpose is MAPE, which computes the mean absolute percentage error between predicted value (GB prediction) and true value (CFD simulation). Instead of running a CFD model, a well-trained GB model can provide high accuracy (84%) $CIOI_v$ predictions for generic urban layouts in real time. This is acceptable and extremely important to support design making in the early design stage.

Table 8
Statistical performance of GB model on the final test set (i=3980, n=20).

R^2_{adj}	RMSE	MAE	MAPE
0.84	0.09	0.06	0.16

Fig. 16 displays the contributions to the model's error deduction, which provides a ranking for each key design variable. With a ratio of 27%, OWR has the highest influence on the $CIOI_v$ value. Thus, opening size is the most efficient design variable for designers to improve indoor ventilation potential. Urban planners and building designers can learn the impact of each design parameter on ventilation potential through sensitivity analysis and further quantify the energy saving potential using natural ventilation.

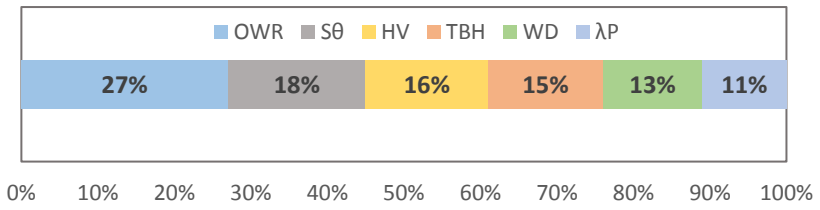


Fig. 16. Contribution of each key design variable to the $CIOI_v$ RMSE deduction.

5. Limitations and future work

This paper has presented the concept of CIOI index for indoor natural ventilation, implemented it using a machine learning prediction model and demonstrated the application on a high-density urban area. The data-driven prediction model shows good performance and the feasibility of supporting early design. However, future work can be improved based on the following limitations.

a) This paper focused on high-density urban areas, where complex airflow behaviors (such as turbulence due to higher roughness, more vortices, backflow) exist. The relationship between design variables and coupled indoor environment is highly non-linear and difficult to describe and explain in generic term. Low density area may not need such complex prediction model.

b) It is assumed that there are eight surrounding blocks around the target building for a generic urban scenario. If for a rare scenario, the target building is located on the boundary of a selected high-density urban area with surrounding blocks less than eight, the CIOI definition is still valid. $A_{out,Ref}$ in Equation 10 will be the area enclosed by the number of adjacent building blocks.

c) The GB prediction model is a data-driven model, which relies on the database quality. This is the common limitation of all machine learning models. Better prediction is expected if more and better data are available for training.

d) This paper assumes that all openings are fully open to achieve maximum cross ventilation potential. If some of the openings are closed, the indoor airflow pattern may change. Partially closed openings will be considered in future work.

e) In this paper, the target building is assumed to be an open-space office building. The building interior layout and obstacles, such as furniture, are ignored in the base case setup. For future work, furniture will be included to study its impact on indoor airflow pattern.

6. Conclusions

Indoor wind environment is highly related to the outdoor wind field. It is therefore necessary to consider outdoor natural ventilation performance for indoor airflow evaluation. This paper has developed an urban-scale coupled indoor and outdoor CFD model to study wind-driven cross ventilation behavior. The coupled CFD model provides both indoor and outdoor wind environments in the same computational domain, which creates the possibility of comparing the wind velocity ratio of indoor and outdoor spaces. Wind tunnel experiment data from the literature is used to validate the base case model. Stochastic simulations are conducted to generate 1,840 paramedic design scenarios by varying the six key design parameters. By

taking advantage of the symmetric feature of the base case model, in a total of 3,840 CFD cases are generated to cover all 16 wind conditions.

A novel ventilation index has been proposed to assess the performance of cross ventilation. The $CIOI_v$ provides the possibility to link indoor velocity to outdoor velocity and shows the ventilation potential, which is not considered in the current natural ventilation guidelines and standards. When $0<\mathrm{CIOI}_v<1$, a higher CIOI_v indicates a smaller velocity loss after the air enters the indoor space, which shows higher building natural ventilation potential.

CFD simulation requires engineering knowledge and expensive computational resource, which creates a barrier for people to use CIOI in the early design stage. To solve this challenge, data-driven CIOI prediction models are developed using machine learning algorithms. Multivariate linear regression (MLR) and gradient boosting (GB) models are training based on the coupled CFD numerical results. Due to smaller gaps among building blocks, the airflow behavior in high-density cities is much more complex compared to that in a low-density urban model. Additionally, by adding the indoor and outdoor interactions makes the relationship between design variables and wind performance are highly non-linear. MLR shows very limited prediction accuracy with MAPE=0.50 and R^2_{adj}=0.29 on the test set. The model training using more advanced GB machine learning algorithm demonstrate much better performance (MAPE=0.16 and R^2_{adj}=0.84). Instead of waiting for hours and days to develop a complicated urban-scale CFD model, the proposed data-driven model can provide acceptable CIOI prediction in real time. This data-driven model is 3~4 orders of magnitude faster than running a CFD simulation. This is very useful to support design decision making. Especially in the early design stage, it may experience several rounds of changes before the design can be finalized. Fast interactions with urban planners and building designers are more helpful for them to recognize the correct trend of achieving better natural ventilation. In the future work, sensitivity analysis of each design variable in the GB model can be conducted to provide a clear understanding of the design impact on ventilation potential and further quantify the energy saving potential using natural ventilation.

Acknowledgements

This work is sponsored by the Ph.D. fellowship at Carnegie Mellon University. The authors would like to thank Professor Nyuk Hien Wong from the National University of Singapore for the illuminating discussions on the concepts of CIOI index. The authors also thank Mr. Cheng Li for his assistance in the Gradient Boosting algorithm.

References

[1] United Nations. Department of Economic and Social Affairs, Population Division. World urbanization prospects: the 2014 revision (Highlight). United Nations Publications; 2014

[2] John Dulac, Marc LaFrance, Nathalie Trudeau and Hirohisa Yamada. Transition to Sustainable Buildings: Strategies and Opportunities to 2050. International Energy Agency; 2013

[3] International Energy Agency and the United Nations Environment Programme. 2018 Global Status Report: towards a zero-emission, efficient and resilient buildings and construction sector; 2018

[4] Luo M, Cao B, Damiens J, Lin B, Zhu Y. Evaluating thermal comfort in mixedmode buildings: a field study in a subtropical climate. Build Environ 2015; 88:46–54

[5] Fisk, W.J. and A. H. Rosenfeld. Estimates of Improved Productivity and Health from Better Indoor Environments. Indoor Air 1997; 7, 158-172.

[6] Wood A, Salib R. Natural ventilation in high-rise office buildings (CTBUH Technical Guides); 2013.
[7] Oropeza-Perez I, Østergaard PA. Energy saving potential of utilizing natural ventilation under warm conditions – a case study of Mexico. Applied Energy 2014; 130:20–32.
[8] Tong Z, Chen Y, Malkawi A, Liu Z, Freeman RB. Energy saving potential of natural ventilation in China: the impact of ambient air pollution. Appl Energy 2016a; 179:660–8
[9] Tong Z, Chen Y, Malkawi A. Defining the influence region in neighborhood-scale CFD simulations for natural ventilation design. Applied Energy 2016b; 182:625-33.
[10] Kato, S., Murakami, S., Mochida, A., Akabayashi, S. I., & Tominaga, Y. Velocity-pressure field of cross ventilation with open windows analyzed by wind tunnel and numerical simulation. Journal of Wind Engineering and Industrial Aerodynamics 1992; 44(1-3), 2575-2586.
[11] Sandberg, M. An alternative view on the theory of cross-ventilation. International journal of ventilation 2004; 2(4),409-418.
[12] Karava, P., Stathopoulos, T. and Athienitis, A. Airflow assessment in cross-ventilated buildings with operable façade elements. Building and Environment 2011; 46(1), 266–279.
[13] Ramponi, R. and Blocken, B. CFD simulation of cross-ventilation for a generic isolated building: impact of computational parameters. Building and Environment 2012; 53, 34–48.
[14] Cheung, J. O., & Liu, C. H. CFD simulations of natural ventilation behaviour in high-rise buildings in regular and staggered arrangements at various spacings. Energy and Buildings 2011; 43(5), 1149-1158.
[15] Larsen, T. S., Nikolopoulos, N., Nikolopoulos, A., Strotos, G., & Nikas, K. S. Characterization and Prediction of the Volume Flow Rate Aerating a Cross Ventilated Building by Means of Experimental Techniques and Numerical Approaches. Energy and Buildings 2011; 43 (6),1371–1381.
[16] Shetabivash, H. Investigation of opening position and shape on the natural cross ventilation. Energy and Buildings 2015; 93, 1-15.
[17] Teppner, R., Langensteiner, B., Meile, W., Brenn, G., & Kerschbaumer, S. Air change rates driven by the flow around and through a building storey with fully open or tilted windows: An experimental and numerical study. Energy and buildings 2014; 80, 570-583.
[18] Van Hooff T, Blocken B, Tominaga Y. On the accuracy of CFD simulations of cross-ventilation flows for a generic isolated building: comparison of RANS, LES and experiments. Building Environ 2017; 114:148–65
[19] Van Hooff, T., & Blocken, B. On the effect of wind direction and urban surroundings on natural ventilation of a large semi-enclosed stadium . Computers & Fluids 2010; 39(7), 1146-1155.
[20] Van Hooff, T, and B Blocken. CFD evaluation of natural ventilation of indoor environments by the concentration decay method: CO2 gas dispersion from a semi-enclosed stadium. Building and Environment 2013; 61,1-17.
[21] Yang, F., Kang, Y., Gao, Y., & Zhong, K. Numerical simulations of the effect of outdoor pollutants on indoor air quality of buildings next to a street canyon. Building and Environment 2015; 87, 10-22.
[22] Buccolieri, R., Sandberg, M., & Di Sabatino, S. City breathability and its link to pollutant concentration distribution within urban-like geometries. Atmospheric Environment 2010; 44(15), 1894-1903.
[23] Hang, Jian, and Yuguo Li. Ventilation Strategy and Air Change Rates in Idealized High-Rise Compact Urban Areas. Building and Environment 2010a; 45 (12), 2754–67.
[24] Hang, Jian, and Yuguo Li. Wind Conditions in Idealized Building Clusters: Macroscopic Simulations Using a Porous Turbulence Model. Boundary-Layer Meteorology 2010b; 136 (1), 129–59.

[25] Lin, M., Hang, J., Li, Y., Luo, Z., & Sandberg, M. Quantitative ventilation assessments of idealized urban canopy layers with various urban layouts and the same building packing density. Building and Environment 2014; 79, 152-167.
[26] Ramponi, R., Blocken, B., Laura, B., & Janssen, W. D. CFD Simulation of Outdoor Ventilation of Generic Urban Configurations with Different Urban Densities and Equal and Unequal Street Widths. Building and Environment 2015; 92,152-66.
[27] Ng, E. Policies and technical guidelines for urban planning of high-density cities – air ventilation assessment (AVA) of Hong Kong. Building and Environment 2009; 44, 1478-1488.
[28] Lee, R. X. Development of estate level outdoor ventilation prediction model for HDB estates in Singapore (Doctoral dissertation). Department of building, National University of Singapore, Singapore; 2013
[29] Ignatius, M., Wong, N. H., & Jusuf, S. K. Urban microclimate analysis with consideration of local ambient temperature, external heat gain, urban ventilation, and outdoor thermal comfort in the tropics. Sustainable Cities and Society 2015; 19, 121-135.
[30] Yuan,C., Norford, L.,Britter,R., Ng,E., A modelling-mapping approach for fine-scale assessment of pedestrian-level wind in high-density cities. Building and Environment 2016; 97,152-165.
[31] Dong, B., Cao, C., & Lee, S. E. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings 2005; 37(5), 545-553.
[32] Kalogirou, S. A. Artificial neural networks in energy applications in buildings. International Journal of Low-Carbon Technologies 2006; 1(3), 201-216.
[33] Yokoyama, R., Wakui, T., & Satake, R. Prediction of energy demands using neural network with model identification by global optimization. Energy Conversion and Management 2009; 50(2), 319-327.
[34] Wong, S. L., Wan, K. K., & Lam, T. N. Artificial neural networks for energy analysis of office buildings with daylighting. Applied Energy 2010; 87(2), 551-557.
[35] Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. Applying support vector machine to predict hourly cooling load in the building. Applied Energy 2009; 86(10), 2249-2256.
[36] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy 2014; 123:168–78
[37] Zhao, J., Lasternas, B., Lam, K. P., Yun, R., & Loftness, V. Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining . Energy and Buildings 2014; 82, 341-355.
[38] Li, X., Wen J., Bai EW. Developing a whole building cooling energy forecasting model for on-line operation optimization using proactive system identification. Applied Energy 2016; 172, 251-263
[39] Fan C., Xiao F., Zhao Y. A short-term building cooling load prediction method using deep learning algorithms. Applied energy 2017; 195, 222-233
[40] Ding, C., & Lam, K. P. Pollution Transport Through Openings Based on Coupled Indoor and Outdoor Interactions. IBPSA-USA Journal 2016a; 6(1).
[41] Architectural Institute of Japan. Guidebook for Practical Applications of CFD to Pedestrian Wind Environment around Buildings. Retrieved from Architectural Institute of Japan: http://www.aij.or.jp/jpn/publish/cfdguide/index_e.htm; 2015.3.20
[42] Yoshie, R., Mochida, A., Tominaga, T., Kataoka, H., Yoshikawa, M. Cross Comparisons of CFD Prediction for Wind Environment at Pedestrian Level around Buildings. Part 1 Comparison of Results

for Flow-field around a High-rise Building Located in Surrounding City Blocks. APCWE-VI . Seoul, Korea; 2005

[43] Franke, J., Hellsten, A., Schlunzen, K. H., & Carissimo, B. The COST 732 Best Practice Guideline for CFD simulation of flows in the urban environment: a summary. International Journal of Environment and Pollution 2011; 44(1-4), 419-427.

[44] Department of Energy. Commercial Reference Buildings . Retrieved from U.S. Department of Energy: http://energy.gov/eere/buildings/commercial-reference-buildings; 2015, 3 20

[45] Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M., & Shirasawa, T. AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings. Journal of wind engineering and industrial aerodynamics 2008; 96(10), 1749-1761.

[46] Ding, C., Lam, K. P., & Wong, N. H. Coupled Natural Ventilation Modeling for Contextual Parametric Design Decision Support. Procedia Engineering 2016b; 169, 264-271.

[47] Yoshie, R., Mochida, A., Tominaga, Y., Kataoka, H., Harimoto, K., Nozu, T., & Shirasawa, T. Cooperative project for CFD prediction of pedestrian wind environment in the Architectural Institute of Japan. Journal of wind engineering and industrial aerodynamics 2007; 95(9), 1551-1578.

[48] Adolphe, L. A simplified model of urban morphology: application to an analysis of the environmental performance of cities. Environment and Planning (B): Planning and Design 2001; 28(2), 183-200.

[49] Yoshida, H. and Omae, M. An approach for analysis of urban morphology: methods to derive morphological properties of city blocks by using an urban landscape model and their interpretations. Computers, Environment and Urban Systems 2005; 29, 223-247.

[50] Golany, G. Urban design morphology and thermal performance. Atmospheric Environment 1996; 30(3), 455-465.

[51] MacDonald, R.W., Griffiths, R.F. and Hall, D.J. An improved method for the estimation of surface roughness of obstacle arrays. Atmospheric Environment 1998; 32(11), 1857-1864.

[52] Raupach, M. Drag and drag partition on rough surfaces. Boundary-layer Meteorology 1992; 60(4), 375-395.

[53] Zhang, A., Gao, C-L. and Zhang, L. Numerical simulation of the wind field around different building arrangements. Journal of Wind Engineering and Industrial Aerodynamics 2005; 93, 891-904.

[54] Oke, T. Street Design and Urban Canopy Layer Climate. Energy and Buildings 1988; 11(1), 103–13.

[55] Grimmond, CSB, and Timothy R Oke. Aerodynamic Properties of Urban Areas Derived from Analysis of Surface Form. Journal of Applied Meteorology 1999; 38 (9), 1262–92.

[56] Vorhies, W. Want to Win Competitions? Pay Attention to Your Ensembles. Retrieved from Data Science Central: http://www.datasciencecentral.com/profiles/blogs/want-to-win-at-kaggle-pay-attention-to-your-ensembles; 2017, 3

[57] Li, C. GB Regressor. Retrieved from Github: https://github.com/cheng-li/pyramid/wiki/GB-Regressor; 2017, 3

[58] Friedman, J. H. Greedy function approximation: a gradient boosting machine. Annals of statistics 2001; 1189-1232.

[59] Gençay, R. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. IEEE Transactions on Neural Networks 2001; 12(4), 726-734.

[60] Aflaki A., Mahyuddin N., Baharum M.R. The influence of single-sided ventilation towards the indoor thermal performance of high-rise residential building: A field study. Energy and Buildings 2016; 126, 146-158

[61] Kosutova K., Van Hooff T., Vanderwel C., Blocken B., Hensen J. Cross-ventilation in a generic isolated building equipped with louvers: Wind-tunnel experiments and CFD simulations. Building and Environment 2019;154, 263-280

[62] Arinami Y., Akabayashi S., Tominaga Y., Sakaguchi J. Performance evaluation of single-sided natural ventilation for generic building using large-eddy simulations: Effect of guide vanes and adjacent obstacles. Building and Environment 2019; 154, 68-80

[63] Liu S., Pan W., Cao Q., Long Z., Jiang Y., Chen Q. CFD simulations of natural cross ventilation through an apartment with modified hourly wind information from a meteorological station. Energy and Buildings 2019; 195, 16-25

[64] N. Ikegaya, S. Hasegawa, A. Hagishima, Time-resolved particle image velocimetry for cross-ventilation flow of generic block sheltered by urban-like block arrays. Building and Environment 2019; 147, 132-145.